

**PATENT APPLICATION**

**METHOD AND SYSTEM FOR CREATING A DOCUMENT  
INTEREST PROFILE**

**Inventor(s):**

Jamey Graham  
136 Sylvian Way  
Los Altos, CA 94022  
a citizen of the UNITED STATES OF AMERICA

**Assignee:**

Ricoh Company, LTD  
3-6 Naka-Magone, 1-Chome  
Ota-Ku, Tokyo 143, Japan  
  
A Japanese Corporation

**Entity:** Large

## **METHOD AND SYSTEM FOR CREATING A DOCUMENT INTEREST PROFILE**

### **CROSS-REFERENCES TO RELATED APPLICATIONS**

This application incorporates by reference the following commonly owned co-pending U.S. Patent Application in its entirety for all purposes:

U.S. Patent Application Serial No. 08/995,616, Entitled , "AUTOMATIC ADAPTIVE DOCUMENT HELP SYSTEM." (attorney docket no. 15358-004200US / ID-CRC-176).

### **BACKGROUND OF THE INVENTION**

The present invention relates to electronic documents and more particularly to method and system for providing interest profiles for electronic documents with features to enhance the experience of reading or using the electronic document.

Increasingly, readers of documents are being called upon to assimilate vast quantities of information in a short period of time. To meet the demands placed upon them, readers find they must read documents "horizontally," rather than "vertically," i.e., they must scan, skim, and browse sections of interest in multiple documents rather than read and analyze a single document from beginning to end.

Documents are becoming more widely available in electronic form. Some documents are available electronically by virtue of their having been created using word processing software. Other electronic documents are accessible via the Internet. Yet others may become available in electronic form by virtue of being scanned in, copied, or faxed. Commonly assigned U.S. Application No. 08/754,721, entitled AUTOMATIC AND TRANSPARENT DOCUMENT ARCHIVING, the contents of which are incorporated herein by reference for all purposes, details a system for generating electronic as well as hardcopy format of documents.

However, the mere availability of documents in electronic form does not assist the reader in confronting the challenges of assimilating information quickly.

Certain tools take advantage of the electronic form documents to assist harried readers. Tools exist to search for documents both on the Internet and locally.

5 Once a document is identified and retrieved, automatic summarization techniques, such as the Reader's Helper<sup>TM</sup>, described in a commonly owned copending U.S. Patent Application No. 08/995,616, entitled AUTOMATIC ADAPTIVE DOCUMENT HELP SYSTEM, the contents of which are incorporated herein by reference for all purposes, helps the reader to find as well as assimilate the information he or she wants more  
10 quickly.

What is needed is a document interest profiling method that helps the reader find the information he or she wants more quickly. The document interest profiling method should be easily personalized, flexible and adaptive as well.

## 15 SUMMARY OF THE INVENTION

According to one embodiment of the present invention, methods and systems for displaying an interest profile for an electronically stored document are provided. Interest profiles provide features that can enhance the experience of reading or using the electronic document. In exemplary embodiments, methods and systems include  
20 one or more interest profile techniques, such as graphical presentations and the like, for browsing or searching documents are provided. The graphical presentation can provide information about content of a document. The invention can be embodied in computer systems that include user input devices, processors, displays, storage and the like.

In an exemplary embodiment, a method for providing an interest profile  
25 includes a variety of steps. A step of accepting from a user input indicating the user's specified concepts of interest can be part of the method. The method can also include analyzing an electronically stored document to identify locations of discussion of the user-specified concept of interest. Techniques for determining locations of concepts of interest can be used with the method, such as keyword counting, Bayesian analysis  
30 techniques described in a commonly owned copending U.S. Patent Application No. 08/995,616, entitled AUTOMATIC ADAPTIVE DOCUMENT HELP SYSTEM, and the like, in various embodiments of the present invention. The method also displays an indication of presence of discussion about the concepts of interest within the

electronically stored document. The indication can provide to a reader approximate positions of portions of the document relevant to the concept of interest.

In some embodiments, the indication can be a graphical representation of the presence of one or more concepts of interest to the reader. The graph can also  
5 indicate a persistence of the concept at various locations within the document by showing a relative amount of discussion of the concepts at various locations within the document. Embodiments can provide graphical presentations including contours, bar charts, scatter plots, and the like.

In another aspect according to the present invention, a method for  
10 providing an interest profile for an electronically stored document comprises a number of steps, such as accepting user input indicating user-specified concepts of interest and analyzing the electronically stored document to identify locations of discussion of the user-specified concepts of interest. The method can include displaying one or more selectable concept indicators corresponding to particular concepts of interest and  
15 accepting user input selecting at least one of these concept indicators. The method can also include displaying an indication of presence of discussion about one or more of the concepts of interest within the electronically stored document. The indication of presence can comprise a contour graph image corresponding to the electronically stored document that indicates a relative presence of discussions of at least one of the concepts of interest  
20 based upon the selections made by the user.

Some embodiments can also include displaying a second indication of presence of the concepts of interest comprising an elongated thumbnail image of all or a portion of electronically stored document in another viewing area of the display. The thumbnail image can have one or more indications of locations of discussion of concepts  
25 of interest based upon the concept or concepts selected by the user. Further, some embodiments will display the second indication juxtaposed to the first indication, enabling the user to visually compare the two indications.

In another aspect according to the present invention, a computer program product for providing an interest profile for an electronically stored document comprises a  
30 computer readable storage medium containing code for performing a variety of tasks. The product can comprise code for accepting user input indicating user-specified concepts of interest and code for analyzing the electronically stored document to identify locations of discussion of user-specified concepts of interest. The program product can

also include code for displaying an indication of presence of discussion about concepts of interest within the electronically stored document. The indication can provide a reader an approximate position of discussion within the electronically stored document.

In another aspect according to the present invention, a system for  
5 providing an interest profile for an electronically stored document can include a memory, a display and a processor interconnected to the memory and the display by a bus. The processor can perform tasks such as accepting user input indicating user-specified concepts of interest and analyzing the electronically stored document to identify locations  
10 of discussion of the user-specified concept of interest. The processor can also display an indication of presence of discussion about the concepts of interest within the electronically stored document on the display. The indication can provide a reader an approximate position of the discussion within the electronically stored document.

Numerous benefits are achieved by way of the present invention over conventional techniques. The present invention can provide an interest profile of a  
15 document for the reader. In many embodiments, the interest profile can be easily personalized to the user. Some embodiments are easier to use than conventional browser user interfaces. Embodiments according to the invention can provide more information to the reader of a document than known techniques. Many embodiments according to the invention will enhance the user's reading or web browsing experience by providing  
20 indication of which portions of documents are likely to be of most interest to the reader. These and other benefits are described throughout the present specification.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

25

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1B illustrate document interest profiles in particular embodiments according to the present invention;

Fig. 2 is an illustration of a system in a particular embodiment according to  
30 the present invention;

Fig. 3 is an illustration of basic subsystems of the system of Fig. 2;

Fig. 4 illustrates a representative screen display in a particular embodiment according to the present invention;

Fig. 5 illustrates a representative screen display in an alternative embodiment according to the present invention;

Figs. 6A-6B illustrate simplified flowcharts of representative process steps in particular embodiments according to the present invention; and

5 Figs. 7A-7D illustrate a representative example of browsing documents in a particular embodiment according to the present invention.

### DESCRIPTION OF THE SPECIFIC EMBODIMENTS

10 The present invention provides interest profiles for electronic documents to enhance the experience of reading or using the document. In exemplary embodiments, methods and systems including a graphical presentation of an interest profile assist the user in browsing or searching documents. The graphical presentation can provide information about content of a document, for example. The invention can be embodied in  
15 computer systems that include user input devices, processors, displays, storage and the like.

Fig. 1A is an illustration of a representative document interest profile image in a particular embodiment according to the present invention. This diagram is merely an example which should not limit the scope of the claims herein. One of  
20 ordinary skill in the art would recognize many other variations, alternatives, and modifications. Fig. 1A illustrates an annotation contour graphical presentation 101 of a document interest profile. Annotation contour 101 presents a profile view of a document based on a persistence measure of a concept within the document. A persistence measure can be a number of annotations ("hits") in the document, for example. Reference may be  
25 had to a commonly owned, co-pending U.S. Patent Application Serial No. 08/995,616, entitled, "AUTOMATIC ADAPTIVE DOCUMENT HELP SYSTEM," incorporated herein by reference in its entirety for all purposes, for a detailed explanation of analyzing a document's content.

Annotation contour 101 can show the concentration, or persistence, of  
30 annotations throughout an entire document, or in portions of the document. Annotation contours can assist users viewing long documents. While annotation contour 101 comprises a line graph format, those of ordinary skill in the art can readily appreciate that other presentation formats, such as bar charts, scatter plots and the like can also be used

without departing from the scope of the present invention. Annotation contour 101 comprises a page pointer 102 that can be dragged to a location on a graphical representation of the document contents 104 by the user using a pointing device, such as a mouse, for example. Responsive to the user's positioning the page pointer to a position in the graphical document representation, the display screen will reposition to display a corresponding location within the document. In this manner, the user can evaluate where the concentration of a concept of interest is within the document and move to this position within the document by sliding the page pointer to a corresponding location on the graphical representation of the document. In a presently preferable embodiment, the distance between the vertical lines of page pointer 102 is approximately 1 page length. However, other lengths can be used without departing from the scope of the present invention.

Fig. 1B is an illustration of another representative document interest profile image in a particular embodiment according to the present invention. This diagram is merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications. Fig. 1B illustrates an annotation contour graphical presentation 103 of a document interest profile. Annotation contour 103 presents a profile view of a document based on a persistence measure of two concepts within the document. Annotation contour 103 comprises a graphical representation 106 corresponding to the persistence of a first concept of interest within the subject document, and a graphical representation 108 corresponding to the persistence of a second concept of interest within the subject document. While annotation contour 103 has been discussed with two concepts of interest, those of ordinary skill in the art can readily appreciate that the method is easily extendible to any number of concepts of interest without departing from the scope of the present invention.

In an alternative embodiment, a single contour can be used to indicate presence or persistence of multiple concepts of interest. In such an embodiment, the contour can display a result of a manipulation of constituent presence or persistence measurements. For example, information about the presence or persistence of two concepts of interest can be added, and a contour can be displayed showing the result of the addition. Other manipulations, such as subtraction of one interest profile from

another, scaling by a scalar or functional factor, and the like are embodiments within the scope of the present invention.

Fig. 2 is an illustration of a representative system suitable for implementing the according to a particular embodiment of the present invention. This diagram is merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications. Fig. 2 depicts but one example of many possible computer types or configurations capable of being used with the present invention. Embodiments according to the invention can be implemented in a single application program such as a browser, or may be implemented as multiple programs in a distributed computing environment, such as a workstation, personal computer or a remote terminal in a client server relationship. Fig. 2 shows computer system 210 including display device 220, display screen 230, cabinet 240, keyboard 250, scanner 260 and mouse 270. Mouse 270 and keyboard 250 are representative "user input devices." Other examples of user input devices are a touch screen, light pen, track ball, data glove and so forth. Fig. 2 is representative of but one type of system for embodying the present invention. It will be readily apparent to one of ordinary skill in the art that many system types and configurations are suitable for use in conjunction with the present invention.

In a preferred embodiment, computer system 210 includes a Pentium® class based computer, running Windows® NT operating system by Microsoft Corporation. However, the method is easily adapted to other operating systems and architectures without departing from the scope of the present invention.

Mouse 270 may have one or more buttons such as buttons 280. Cabinet 240 houses familiar computer components such as disk drives, a processor, storage means, etc. As used in this specification "storage means" includes any storage device used in connection with a computer system such as disk drives, magnetic tape, solid state memory, bubble memory, etc. Cabinet 240 may include additional hardware such as input/output (I/O) interface cards for connecting computer system 210 to external devices such as a scanner 260, external storage, other computers or additional peripherals.

Fig. 3 is an illustration of basic subsystems in computer system 210 of Fig. 2. In Fig. 3, subsystems are represented by blocks such as central processor 300, system memory 310, etc. This diagram is merely an example which should not limit the scope of



the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications. The subsystems are interconnected via a system bus 320. Additional subsystems such as a printer, keyboard, fixed disk and others are shown.

5      Peripherals and input/output (I/O) devices can be connected to the computer system by any number of means known in the art, such as serial port 330. For example, serial port 330 can be used to connect the computer system to a modem, a mouse input device, or a scanner. The interconnection via system bus 320 allows central processor 300 to communicate with each subsystem and to control the execution of instructions from system memory 310 or the fixed disk, as well as the exchange of information between  
10      subsystems. Other arrangements of subsystems and interconnections are readily achievable by those of ordinary skill in the art. System Memory 310, and the fixed disk are examples of tangible media for storage of computer programs, other types of tangible media include floppy disks, removable hard disks, optical storage media such as CD-ROMS and bar codes, and semiconductor memories such as flash memory, read-only-  
15      memories (ROM), and battery backed memory.

Fig. 4 illustrates a representative browser user interface 401 having a document interest profile display 402 which can be displayed on display screen 230 of Fig. 2, in a particular embodiment according to the present invention. This diagram is merely an example which should not limit the scope of the claims herein. One of  
20      ordinary skill in the art would recognize many other variations, alternatives, and modifications. Fig. 4 illustrates an example use of an annotation contour 402 in conjunction with a find command box 404 for searching through the current document of interest, such as a page from the world wide web, for example, for information about one or more concepts of interest. As shown in Fig. 4, a user has entered a topic, "paper" in a  
25      field within the find command box 404. By clicking on the "find next" button, the user can move through the document from one instance of the word "paper" to the next. Additionally, annotation contour 402 indicates a persistence of the term "paper" within the document. A page pointer 406 at the top of annotation contour 402, can provide an indication of the relative position of the current page displayed to the user in the display  
30      window.

Fig. 5 illustrates another representative browser user interface having a document interest profile display 502 which can be displayed on display screen 230 of Fig. 2, in a particular embodiment according to the present invention. This diagram is

merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications. Fig. 5 illustrates a user interface 501 for viewing an annotated document online. Annotations can be added to a document in a variety of ways, such as described in a commonly owned copending U.S. Patent Application No. 08/995,616, entitled AUTOMATIC ADAPTIVE DOCUMENT HELP SYSTEM, which is incorporated herein by reference in its entirety for all purposes. A first viewing area 502 in user interface 501 shows a section of an electronic document. Using a scroll bar 504, or in other ways, the user may scroll the displayed section through the electronic document.

Some embodiments can include a series of concept indicators 506 that permit the user to identify which concepts of interest are to be noted in the document. A sensitivity control 508 permits the user to select the degree of sensitivity to apply in identifying potential locations of relevant discussion. At low sensitivity, more locations will be denoted as being relevant, even though some may not be of any actual interest. At high sensitivity, most all denoted locations will in fact be relevant but some other relevant locations may be missed. Many embodiments provide a percentage giving the relevance of the currently viewed document to the concept for each concept name appearing by one of selectable concept indicators 506. These relevance levels offer a quick assessment of the relevance of the document to the selected concepts.

In many embodiments, an elongated thumbnail image 514 representing the contents of the document is provided in a second viewing area 515. Thumbnail image 514 depicts a representation of the contents of the document, the current location of the reader, i.e., the location of the text displayed in section 502, using a sliding window 522, as well as annotations corresponding to locations of discussion of concepts of interest to the reader, such as annotation 530a that corresponds to annotation 530b in the document displayed in viewing area 502. An annotation contour 520 can be displayed along with elongated thumbnail image 514. Annotation contour 520 can provide an indication of the presence of one or more concepts of interest to the reader within the document. A second sliding window 524 provides a counter part to sliding window 522. Counterpart sliding window 524 can be positioned along the annotation contour 520, to enable the user to find discussions of the topics of interest within the document.

In some embodiments, one or more navigation tools can be found on a navigation toolbar 516. Miscellaneous annotation tools can be found on an annotation

toolbar 518. The annotation tools on annotation toolbar 518 facilitate navigation through a collection of documents. Annotations may be added to the document text to denote relevance to user-selected concepts of interest.

Fig. 6A illustrates a flowchart 601 of simplified process steps in a particular representative embodiment according to the invention for creating a document interest profile. This diagram is merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications. The method is preferably embodied as software processes executing in a computer system that includes a user input device coupled to a processor, a display and a memory. In a step 602, user input indicating user-specified concepts of interest is accepted. Next, in a step 604, the electronic document is analyzed in order to identify locations of discussion of the concepts of interest to the user specified in step 602. Then, in a step 606, the electronic document is displayed, with an indication to the reader of the presence of the concepts of interest within the document. The indication can take the form of a contour graph that indicates the relative strength of the concept of interest at various points within the document. In some embodiments, a pointer associated with the graph can provide an indication to a reader of an approximate position of a currently displayed portion of the document, using a 'sliding window' style indication.

Fig. 6B illustrates a flowchart 603 of simplified process steps in another representative embodiment according to the invention for creating a document interest profile. This diagram is merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications. Fig. 6B illustrates a step 612, wherein user input indicating user-specified concepts of interest is accepted. Next, in a step 614, the electronic document is analyzed in order to identify locations of discussion of the concepts of interest to the user specified in step 612. In a step 615, input from the user selecting one or more concepts of interest is accepted. Input can be received from the user clicking on icons with the mouse, or the like. Then, in a step 616, the electronic document is displayed, with an indication to the reader of the presence of the concepts of interest selected by the user in step 615. The indication can take the form of a contour graph that indicates the relative strength of the concept of interest at various points within the document. In some embodiments, a pointer associated with the graph can provide an

indication to a reader of an approximate position of a currently displayed portion of the document, using a 'sliding window' style indication.

Some embodiments will also include a document thumbnail image to provide summary information, and the like.

5 Figs. 7A-7D illustrate representative screens in examples of particular embodiments according to the present invention. In the examples of Figs. 7A-7D, a user is using an exemplary embodiment to learn about wearable computers. Other references about these topics, as well as other topics can be viewed using various embodiments according to the present invention. Thus, the screens of Figs. 7A-7D are intended to  
10 exemplary and not limiting.

Fig. 7A illustrates a representative screen in an example search or browsing of a web based document in a particular embodiment according to the present invention. This diagram is merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations,  
15 alternatives, and modifications. Fig. 7A illustrates a browser user interface 701 for viewing an annotated document online. A first viewing area 702 in user interface 701 shows a section of an electronic document. A scroll bar 704 provides the user with the ability to scroll the displayed section through the electronic document. A plurality of selectable concept indicators 706 permit the user to identify which concepts of interest are  
20 to be noted in the document. A percentage gives the relevance of the currently viewed document to the concept for each concept name appearing by one of selectable concept indicators 706. These relevance levels offer a quick assessment of the relevance of the document to the selected concepts. An annotation 708 indicates the presence of the word "wearable" within the body of the document. "Wearable" is a concept of interest denoted  
25 by concept of interest indicator 710.

Annotation contour 712 provides an "activity indication" of the locations of the annotations, such as annotation 708, within the document. The graph of annotation contour 712 indicates a relative number of annotations throughout the document. Other types of indicators, such as bar graphs, scatter plots and the like can also be used to  
30 provide information about the location and number of annotations within the document in other embodiments according to the present invention. Page pointer 714 indicates the relative position of the currently displayed page of viewing area 702 within the document. The user can move page pointer 714 along annotation contour 712 in order to position the

document in the first viewing area 702 to a portion of the document having a high concentration of instances of annotations.

Fig. 7B illustrates another representative screen in the example search or browsing of a web based document in a particular embodiment according to the present invention. This diagram is merely an example which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives and modifications. Fig. 7B illustrates browser 701 wherein a user has repositioned page pointer 714 to a portion of the document having relatively many instances of annotations for a particular concept. Display viewing area 702 now depicts a portion of the document corresponding to the placement of the page pointer by the user. In this way, the user can quickly locate portions of a large document having material about concepts of interest to the user by sliding the page pointer to an apparent "hot spot" within the document.

Fig. 7C illustrates a further representative screen in an example search or browsing of a web based document in a particular embodiment according to the present invention. This diagram is merely an example, which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives and modifications. Fig. 7C illustrates browser 701 wherein the user has selected an individual concept, "wearable" as being of interest. The user selects a concept indicator 706a corresponding to the concept "wearable" from among the selectable concept of interest indicators 706. Annotation contour 712 depicts the relative occurrence of the selected concept, "wearable" within the document.

Fig. 7D illustrates a yet further representative screen in an example search or browsing of a web based document in a particular embodiment according to the present invention. This diagram is merely an example, which should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, alternatives and modifications. Fig. 7D illustrates browser 701 wherein the user has selected two concepts, "wearable," and "design" as being of interest. The user selects a concept indicator 706a corresponding to the concept "wearable," and a concept indicator 706b corresponding to the concept "design" from among the selectable concept of interest indicators 706. Annotation contour 712 depicts the relative occurrence of the selected concept, "wearable" within the document.

In conclusion, the present invention provides for a method of providing an interest profile for documents. In the foregoing specification, the invention has been described with reference to a specific exemplary embodiments thereof. Many changes or modifications are readily envisioned. For example, changing the size or arrangement of the document interest profile image in the display, changing the appearance and features of document interest profile image by using different graphing methods and other types of depictions; adding audio effects; adding audio memos describing the contents of the document, among other changes, are included within other embodiments of the present invention.

10 The specification and drawings are, accordingly, to be regarded in an illustrative rather than in a restrictive sense. It will, however, be evident that various modifications and changes may be made thereunto without departing from the broader spirit and scope of the invention as set forth in the claims.